l. 21 No. 6 (New Series No. 84) June 1961

ANALYSIS

THE UNIVERSITY OF MICHIGAN
JUN 1961
PERIODICAL READING ROOM

Acting Editor:
P. T. GEACH
with the advice of
A. J. Ayer
R. B. Braithwaite
Herbert Dingle
A. E. Duncan-Jones
C. A. Mace
A. M. MacIver
H. H. Price

CONTENTS

Malcolm and the Use of Words
LEONARD LINSKY

Roundabout the Runabout Inference-Ticket
J. T. STEVENSON

Plausible Implication
NICHOLAS RESCHER

Wittgenstein and Logical Compulsion
CHARLES S. CHIHARA

Paradoxes of Self-Deception
JOHN CANFIELD and PATRICK McNALLY

THREE SHILLINGS AND SIXPENCE NET

ASIL BLACKWELL · BROAD STREET · OXFORD

ANALYSIS 21.6 JUNE 1961

MALCOLM AND THE USE OF WORDS

By LEONARD LINSKY

1. PART of the strangeness of Malcolm's question ' Can a man judge that he is asleep? ' comes from the odd use of the word ' judge ' here.[1] This is a philosopher's use of that word, and not in accordance with the ordinary use at all. To say that I judge someone to be asleep implies that I am not certain, that I don't actually know that that person is asleep. It implies further that I have thought to some extent about the matter and have then concluded whatever I judge to be the case. That is why it would sound so strange for me to say that I judged that my head ached. In such a case, I claim, we use the word ' judge ' partly to indicate a lack of certainty; and as regards my headaches such hesitation is totally out of place—how can I be hesitant about whether my head aches? Further, what is there to reflect about in Malcolm's case? Can we imagine someone deciding that he is *probably* asleep?

But in Malcolm's use of it the word ' judge ' is really meant as a stand-in for each of a group of other words such as ' decide ', ' suppose ', ' wonder ', ' assert ', ' imply '. Nobody can decide that, suppose that, wonder whether, assert that, or imply that he is asleep. The possibility of any of these things implies the meaningfulness of the assertion (claim) that I am asleep; and Malcolm is arguing that the words ' I am asleep ' are not used in our language to express any thought at all.

I do not think that it requires an argument to support Malcolm's claim. It seems obvious to me that were we to hear anyone utter the words ' I am asleep ', we should never think he supposed that he was asleep. This is not to say that if someone were to say these words, in certain circumstances, he would fail to convey anything. If someone raps on my door in the early morning, I may indicate to him that I don't wish to be disturbed (and *why* I don't wish to be disturbed) by calling out ' Go away, I'm asleep '. I think, however, that I would be much more likely to say ' I'm *still* sleeping ' or, if I were being strict in my use of words, ' I'm still trying to sleep '. In any case, Malcolm is right in maintaining that the first-person sentence ' I am asleep ', if it is said at all (except as a logical joke), cannot be regarded as having a completely homogeneous meaning with the third-person form of the same sentence, ' He is asleep '.

2. Malcolm, however, does argue for his claim, and the argument which he produces contains an interesting mistake.[2] Malcolm asks the

[1] N. Malcolm, *Dreaming*, Routledge & Kegan Paul, London, 1959, Chapter III.
[2] *Ibid.*, pp. 9–12.

question: 'How could someone be taught the supposed correct use of the sentence 'I am asleep' to describe his own state?'. Malcolm's supposition here is that if this sentence can be so used in order "to describe one's own state", there must be some way in which we can learn to use the sentence to this end. For, of course, we learn the use of our language from those who already have mastered it. Now, Malcolm observes, a sleeping man cannot be taught the use of the words 'I am asleep' by means of an *ostensive* definition. An ostensive definition could only teach one the meaning of the words 'He is asleep', but never the use of the words 'I am asleep'. Of course, the whole idea of a person while asleep being taught anything by another person is strange.

Further, Malcolm asks, how could one ever verify that another person *understands* how to use these words to describe his own state? We would have to determine that he applies these words to himself at the right time. That is, for the most part, when he says these words what he says should be true. For if always, or usually, when he said these words they were false, one would have good reason to suppose that the speaker did not understand the words in the required sense. "Using a sentence *correctly* and using it to make a *true* statement are different concepts, but the former depends upon the latter in the same sense that the concept of an order depends on the concept of obeying an order."[3]

Suppose that we try to imagine a primitive tribe such that the people of this tribe constantly gave orders to each other, which they never or hardly ever obeyed! This is obviously absurd; for what reason would we have for calling this ceremony (practice) the practice of giving orders? In the same way, it is Malcolm's contention, there is a logical connection between the concept of a correct use of a sentence and the concept of its use to state what is true. If a person never or hardly ever used a sentence to say what is true, what reason should we have for saying that that person had mastered the *correct* use of that sentence?

But can we determine that when a speaker uses the words 'I am asleep', for the most part he uses them to describe his present state correctly? Can we determine that he even does this once? The answer is 'no'; for in order to know that when a man says 'I am asleep' he gives a true description of his own state, one would have to suppose both that he was asleep and that he was aware of what he said. Now these requirements are mutually contradictory, because whatever showed that he was aware of what he said would show at the same time that he was not asleep.

3. Though I believe that there are good elements in this argument

[3] *Ibid.*, pp. 9–10.

of Malcolm's, it also seems to me to involve suppositions which are not correct. Malcolm continually speaks about our *use* of sentences. He speaks of a correct and an incorrect use of the sentence 'I am asleep' to describe one's present state. Now in ordinary language, we do not speak about the use of sentences at all, but about the use of *words* which compose the sentences in question.[4] We can use a word correctly or incorrectly, but we cannot use a sentence correctly or incorrectly. Malcolm's point becomes especially difficult to accept because he insists on distinguishing the *correct* use of a sentence from its use to make a *true* statement.

I might, of course, ask a person what he meant by a certain sentence, and I might ask him how he was using a word, but I would never ask him how he was using a sentence; and surely what Malcolm means to discuss, in his talk about the use of a sentence, is the meaning of that sentence. What I learn in learning language is the correct use of words; part of what I learn, in learning this, is how to combine these words into meaningful sentences. If I master the correct use of words (know the correct use), I can combine them into meaningful sentences; and it is possible for both speaker and hearer to understand these sentences even on the *first* occasion of their utterance. Therefore it is wrong to say, as Malcolm does, that I can only know that a speaker understands the meaning of a sentence if I know that, for the most part, when he says that sentence he says it in the course of making a *true* assertion. For how could I know *that* about a speaker, if I have never heard him say that sentence before at all? Surely there are other things about a person's verbal behavour, besides his saying what is true in saying a sentence, which would justify my belief that this person understands the sentence.

On Malcolm's view, a sentence by the utterance of which a true statement has *never* been made is one that we can never know another person can use significantly, i.e. with understanding. But this is plainly false. Take for example the sentence 'I see a ghost'. No one has ever made a true assertion by uttering this sentence; there are no ghosts. But this certainly does not alter the fact that many people know what it means, and that we know they know what it means.

University of Illinois

[4] This distinction between the use of words and the use of sentences is due to Prof. Gilbert Ryle. See his article 'Ordinary Language', *Philosophical Review* 62 (1953), pp. 178ff.

ROUNDABOUT THE RUNABOUT INFERENCE-TICKET

By J. T. STEVENSON

IN HIS article "The Runabout Inference-Ticket" Professor A. N. Prior tries to show that there is an absurdity derivable from the theory "... that there are inferences whose validity arises solely from the meanings of certain expressions occurring in them."[1] For accounts of the theory other than his own Prior refers us to the writings of K. R. Popper, W. Kneale, P. F. Strawson, and R. M. Hare. I shall not be concerned to determine whether he has accurately represented their versions of the theory (although I think this doubtful for example in the case of Popper), because Prior's interpretation is itself intrinsically interesting. Prior's argument strongly suggests that there *is* something wrong with the theory, as he presents it, but, unfortunately, he does not show us *what* is wrong with it. I wish to show (1) exactly where the theory, *as stated by Prior*, goes wrong, and (2) that the theory can be stated in such a way as to be quite sound.

According to the theory in question, the inference 'Grass is green and the sky is blue, therefore grass is green' is an analytically valid inference solely in virtue of the meaning of the word 'and'. The presumed analyticity of the inference is exhibited by the following statement of the meaning of the word 'and': "... (i) from any pair of statements P and Q we can infer the statement formed by joining P to Q by 'and' ... (ii) from any conjunctive statement P-and-Q we can infer P, and (iii) from P-and-Q we can always infer Q."[2]

Prior attempts to reduce the foregoing theory to absurdity by introducing a new connective 'tonk' and giving it a meaning in the way suggested by the theory. The complete meaning of 'tonk' is: "(i) from any statement P we can infer any statement formed by joining P to any statement Q by 'tonk' ..., and ... (ii) from any 'contonktive' statement P-tonk-Q we can infer the contained statement statement Q."[3] He then shows that the following inference is valid in virtue of the meaning of 'tonk':

2 and 2 are 4.
Therefore, 2 and 2 are 4 tonk 2 and 2 are 5.
Therefore, 2 and 2 are 5.[4]

Prior does not say, but seems to imply, this: Since the theory allows to deduce, "in an analytically valid way", a patently false statement from a patently true one, there must be something radically wrong with the theory.

[1] ANALYSIS 21.2, December 1960.
[2] *Ibid.* [3] *Ibid.* [4] *Ibid.*

In order to understand what has happened here, it is essential to notice that the theory requires us to give the meaning of logical connectives in terms of *rules*. These rules are permissive: I take it that the force of 'we can infer', as it occurs in the foregoing definitions, is the same as 'we *may* infer' or 'we are *allowed* or *permitted* to infer'. If 'we can infer' were taken to mean the same as 'we may *validly* infer', some of the things I shall say would have to be modified. But, in this case, if 'valid' were used in its ordinary sense (namely, such as to lead from truth only to truth and never to falsehood), Prior's definition of 'tonk' would become radically incoherent, indeed self-contradictory, and his argument trivially unsound and hence uninteresting. I shall take the more interesting and more usual interpretation that rules of inference are simply permissive. Granted this, we can now consider two important insights and one serious error in the theory.

The first insight concerns the meaning of logical connectives: the way in which we can express the meaning of connectives must be different from the way in which we express the meanings of non-logical words. In the first place, leaving aside Platonism, connectives are not used to denote, and hence the sort of semantical properties they have will be different from those of non-logical words. Second, logical terms are syncategorematic or incomplete symbols; they have no meaning in isolation. Since the most distinctive feature of a logical term is its syntactical properties, we can explain its meaning to someone unfamiliar with it by exhibiting how these syntactical properties affect the contexts in which the connective in question occurs. And a very convenient way to do this is to give the permissive rules governing the inferences we can make using the connective.

The second insight is that we ordinarily justify (i.e., validate) inferences by appealing to a permissive rule. If you question my inference 'If I don't leave in five minutes, I shall be late, and I am not going to leave in five minutes; so I shall be late', I justify it by appealing to the permissive rule *modus ponens*.

The serious error in the theory consists in combining these two insights in an unfortunate way. It is assumed that we can *completely* justify an inference by appealing to the meaning of a logical connective as stated in permissive rules. If this were so, we could, as Prior shows, justify any inference whatsoever by defining a logical connective in terms of permissive rules in such a way that we would be allowed to pass from true premises to a false conclusion.

The crucial point to be noted is this: in order to *completely* justify an inference we must appeal to a *sound* rule of inference. A complete justification of an inference has two parts: we must first *validate* the inference by subsuming it under a rule, and secondly we must *vindicate*

the rule itself by showing that it is a sound rule.[5] A deductive rule is sound if and only if it permits only valid inferences, an inference being valid in this sense if and only if it is such that when the premises are true the conclusion must be true. The difficulty in our theory, then, is that it does not prevent us from defining connectives in terms of unsound permissive rules. Since no attempt is made to vindicate the rules used in the definitions, the definitions do not, by themselves, provide a complete justification of our inferences.

I now turn to the problem of stating the theory in such a way that it avoids the above difficulty. Basically, the theory states that certain inferences are completely justified solely in virtue of the meanings assigned to certain logical connectives. Since giving the meaning of a logical connective consists in giving its syntactical properties, we must show that, given a statement of the syntactical properties of a connective, the soundness of certain rules of inference can be demonstrated. There is no difficulty in doing this; it can be done, indeed, it has been done, for many different connectives, and there is no need to go into details here.[6]

To be more precise, two qualifications should be made. First: the syntactical properties of a connective include both its formation and transformation properties, although here only its transformation properties are considered. Second: we can exhibit the transformation properties of a sentence connective in a calculus by making a value-table for it either so that the calculus remains uninterpreted, or so that it becomes interpreted. In the former case, we might use some arbitrary symbols for the values (say 0 and 1), and deal with pure syntactical properties. In the latter case, we use truth and falsity as values; and, since truth is a semantical notion, the calculus becomes to some extent interpreted, and we are no longer dealing with pure syntactical properties. For answering questions of soundness the latter method is the one which must be used; but for convenience I continue to speak simply of syntactical properties.

In a formal calculus we can state the syntactical properties of, say, a truth-functional binary sentence connective ' o ', by stating, in the meta-language, the way in which the truth-value of the well-formed formula ' poq ' is a function of (all possible combinations of) the truth-values of the components ' p ' and ' q '. We can then deduce from these statements, in a very rigorous way, a meta-theorem of the calculus (again stated in the meta-language) to the effect that such-and-such permissive rules are sound, i.e., lead from truths only to truths and never to falsehoods. Since from a statement of the meaning of a connective we can derive demonstrably *sound* permissive rules of inference governing that

[5] The distinction between validation and vindication is due to H. Feigl. See " De Principiis non Disputandum . . . ? " in *Philosophical Analysis*, ed. Max Black (Cornell University Press, 1950).

[6] See any standard text, e.g., Church's *Introduction to Mathematical Logic*.

connective, we may say that certain inferences are *completely* justified solely in virtue of the meanings of certain expressions occurring in them.

The important difference between the theory of analytic validity as it should be stated and as Prior stated it lies in the fact that he gives the meanings of connectives in terms of permissive rules, whereas they should be stated in terms of truth-function statements in a meta-language. The theory of analytic validity does *not* require that the meanings of connectives be given in terms of rules; as we have seen, to do so is to leave open the question of complete justification. What the correct theory of analytic validity *does* require is that the meanings of connectives be given in terms of statements of syntactical properties. When this is done the soundness of certain rules of inference is demonstrable, and thus inferences can be completely justified by appealing to the meanings of connectives. Using the latter method we block the introduction of a connective like Prior's ' tonk '. This can be seen as follows.

Consider these two truth-tables which exhibit in a graphic way the syntactical properties of two binary sentence connectives ' o ' and ' § '.

A.

p	q	poq
T	T	T
T	F	T
F	T	F
F	F	F

RA: p ∴ poq

B.

p	q	p§q
T	T	T
T	F	F
F	T	T
F	F	F

RB: p§q ∴ q

From A it can be seen, intuitively, that the syntactical properties of ' o ' permit us to demonstrate that the permissive rule RA, namely, from ' p ' you may infer ' poq ', is sound; and with a properly formulated statement of these syntactical properties it can be rigorously demonstrated to be sound. Similarly, from B it can be seen that the rule RB, namely, from ' p§q ' you may infer ' q ', is a sound rule. Prior's connective ' tonk ' is governed by two rules like RA and RB. The syntactical properties of ' tonk ', then, must be a combination of the syntactical properties of ' o ' and ' § '; and in order to demonstrate the joint soundness of the rules for ' tonk ', we would have to construct a truth-table combining all the features of A and B. But it is obvious that this would involve ascribing contradictory syntactical properties to ' tonk '. This being so, it would be impossible to state consistently the meaning of ' tonk ' in the manner I have suggested; and hence impossible to completely justify the inference from ' 2 and 2 are 4 ' to ' 2 and 2 are 5 '. One could, of course, as Prior has done, state the meaning of ' tonk ' in terms of rules and in this way justify (i.e., validate) ' 2 and 2 are 4, therefore, 2 and 2 are 5 ', but this would not *completely* justify the inference, for it would leave open the question as to the *vindication* of

the inference. And, of course, by definition it could never be vindicated, for it leads from an obvious truth to an obvious falsehood. I conclude, then, that there is nothing wrong with the theory of analytic validity when properly stated.[7]

Brown University

PLAUSIBLE IMPLICATION

By NICHOLAS RESCHER

I

I PROPOSE to examine the formal logic of an implication-relationship which has never, to my knowledge, been studied; a relationship which I shall designate as *plausible implication*. The conception which lies at the basis of this relationship is at once simple and natural. We shall say that a proposition p *plausibly implies* the proposition q if there is some proposition r which meets the conditions that: (i) r itself is *likely* (or probable), and (ii) the conjunction of p with r entails (strictly implies) q. In short, if '$\rightarrow$' represents entailment (strict implication) and 'L' represents the modality "is likely" as applicable to propositions, then we may formally characterize plausible implication, to be symbolized as '$\Rightarrow$', by the definition:

(D1.1) '$p \Rightarrow q$' = Df '$(\exists r)[L(r) \,\&\, ([p \,\&\, r] \rightarrow q)]$'

In accordance with this definition, one proposition plausibly implies another when the conjunction of the former with some suitable proposition which is *likely* yields the latter as an entailed consequence.

To be sure, there are a great many ways of putting an exact construction on the common-language locution 'If p, then quite likely q'.[1] The implication-relationship '$\Rightarrow$' given by the definition (D1.1) clearly represents one, albeit only one, of these. It is this particular relationship which is here designated, for brevity and convenience, as 'plausible implication'. Correspondingly, '$p \Rightarrow q$' may be read as 'If p then it is plausible that q' or more compactly as 'p plausibly implies q'. The present paper will be concerned to examine the formal logic of this

[7] I should like to acknowledge the benefit I have had of a number of stimulating discussions with Wesley C. Salmon on this topic. He is not, of course, responsible for any errors I may have made.

[1] For example, the various ways of construing the conception 'p constitutes good evidence for q' could also serve here. For some of the possibilities that lie in this direction see the writer's paper on 'A Theory of Evidence', *Philosophy of Science*, vol. 25 (1958), pp. 83–94. The cognate conception of a modality of "conditional likelihood" is discussed in *idem*, 'A Contribution to Modal Logic', *The Review of Metaphysics*, vol. 12 (1958), pp. 186–199.

concept. Such a discussion may be of interest in that the logical theory of such essentially *inductive* implication relationships is a very much underdeveloped branch of logic.

II

It is appropriate to note at the outset that there is a significant formal analogy between plausible implication, '$\Rightarrow$', and material implication, '$\supset$'. As John Myhill has shown, in any of the standard systems of strict implication, we have the result that $p \supset q$ is equivalent (interdeducible) with $(\exists r)[r \,\&\, ([p \,\&\, r] \rightarrow q)]$.[2] Thus p implies q *materially* (rather than *plausibly*) when the conjunction of p with some *true* (rather than *likely*) proposition strictly implies q. The formal analogy of these two modes of implication is thus evident.

It must be observed, however, that neither '$p \rightarrow L(p)$' not its converse, '$L(p) \rightarrow p$' are valid theses (see section III below). Thus the foregoing analogy between '$\Rightarrow$' and '$\supset$' does not constitute a basis for concluding that one of these relationships embraces the other. As a result, neither

(N2.1) $\quad (p \supset q) \rightarrow (p \Rightarrow q)$

nor its converse,

(N2.2) $\quad (p \Rightarrow q) \rightarrow (p \supset q)$

are valid theses.

III

Before entering upon a more detailed study of the logical theory of the relationship of plausible implication, it will be helpful to pass in review the logical characteristics of the notions employed in the definition of this conception, namely the strict implication relationship '$\rightarrow$' and the likelihood modality 'L'. As regards the former, we shall assume that the mode of strict implication relation here in question is governed by the rule:

(R3.1) Whenever a proposition q is deducible from a proposition p (by the " usual " rules of logical inference), we have: $p \rightarrow q$.

Moreover, we shall assume that '$\rightarrow$' is subject to the following principles:

(I1) $[(p \rightarrow q) \,\&\, (q \rightarrow r)] \rightarrow (p \rightarrow r)$

(I2) $[(p \,\&\, q) \rightarrow r] \rightarrow [(p \,\&\, \sim r) \rightarrow \sim q]$

(I3) $[(p \rightarrow q) \,\&\, (p \rightarrow r)] \rightarrow [p \rightarrow (q \,\&\, r)]$

In summary, our system of strict implication for '$\rightarrow$' may be thought of as a somewhat strengthened version of Lewis' system S3.[3]

[2] " On the Interpretation of the Sign '$\supset$' ", *The Journal of Symbolic Logic* vol. 18 (1953), pp. 60–62.

[3] For this system, and some cognates, see Appendix II of C. I. Lewis and C. H. Langford, *Symbolic Logic* (N.Y., 1932).

As regards the likelihood modality ' L ', it suffices to note that the principles (L1)—(L5) *do*, while the princples (N3.1)—(N3.3) *do not* obtain.[4]

(L1) $(\sim p \rightarrow p) \rightarrow L(p)$
(L2) $(p \rightarrow \sim p) \rightarrow \sim L(p)$
(L3) $L(\sim p) \rightarrow \sim L(p)$
(L4) $[L(p) \ \& \ (p \rightarrow q)] \rightarrow L(q)$
(L5) $[L(p) \ \& \ ([p \ \& \ q] \rightarrow \sim q)] \rightarrow \sim L(q)$

(N3.1) $p \rightarrow L(p)$
(N3.2) $L(p) \rightarrow p$
(N3.3) $[L(p) \ \& \ L(q)] \rightarrow L(p \ \& \ q)$

Note that in view of the failure of (N3.1) and (N3.2) to characterize ' L ', propositions of the type $p \ \& \sim L(p)$ and $\sim p \ \& \ L(p)$ may be true in appropriate instances.

IV

One of the most significant characteristics of the implication relationship '$\Rightarrow$' is that it is subject to the rule:

(R4.1) Whenever ' $p \rightarrow q$ ' obtains, we also have ' $p \Rightarrow q$ '.

The validity of this rule is easily demonstrated. By the definition (D1.1), we have that ' $p \Rightarrow q$ ' amounts to ' $(\exists r)[L(r) \ \& \ ([p \ \& \ r] \rightarrow q)]$ '. Now assume ' $p \rightarrow q$ '. Then, since ' $(p \ \& \ r) \rightarrow p$ ' obtains by (R3.1), we have, by (I1), for any r, $(p \ \& \ r) \rightarrow q$. Thus ' $p \Rightarrow q$ ' comes to amount to ' $(\exists r) \ L(r)$ ', which is trivially true.

In view of (R4.1) we immediately obtain the following theorems:

(T4.1) $p \Rightarrow p$
(T4.2) $p \Rightarrow \sim \sim p$
(T4.3) $p \Rightarrow (p \ \& \ p)$
(T4.4) $p \Rightarrow (p \vee q)$
(T4.5) $(p \ \& \ q) \Rightarrow p$
(T4.6) $(p \vee q) \Rightarrow (q \vee p)$
(T4.7) $(p \ \& \ q) \Rightarrow (q \ \& \ p)$
(T4.8) $[p \ \& \ (q \ \& \ r)] \Rightarrow [q \ \& \ (p \ \& \ r)]$
(T4.9) $(p \rightarrow \sim p) \Rightarrow \sim p$

And of course a limitless number of similar relationships will govern '$\Rightarrow$' in consequence of (R4.1).

It seems appropriate to list a few further theorems relating to '$\Rightarrow$'. The justification of the theorems now to be listed, unlike those given

[4] For a more detailed treatment of the concept of a likelihood modality see the writer's paper 'A Contribution to Modal Logic' as cited in footnote 1 above. However, in this paper the likelihood modality is symbolized by ' o ' and not by ' L ', which plays another role.

above, does not fall immediately under rule (R4.1). However the proofs of these theorems are so straightforward that we may omit their presentation.

(T4.10) $(p \Rightarrow q) \rightarrow (\sim q \Rightarrow \sim p)$

(T4.11) $(p \Rightarrow \sim p) \rightarrow (p \Rightarrow q)$

(T4.12) $(\sim q \Rightarrow q) \rightarrow (p \Rightarrow q)$

In view of the fact that '$\Rightarrow$' is governed by the various theorems set forth above, this implication-relationship goes a long way towards exhibiting the characteristics of a strict implication. For example, '$\Rightarrow$' has already been shown to satisfy six of the eight axioms of Lewis' system of strict implication S3.[5] However, as we shall shortly see, there are certain characteristics of '$\Rightarrow$' which utterly disqualify it as a mode of strict implication.

V

Several results which link the implication relationship '$\Rightarrow$' with the likelihood modality 'L' deserve to be noted. Two of these may be characterized as "Paradoxes of Plausible Implication" (by analogy with cognate results for material and for strict implication). These are the following theorems:

(T5.1) $L(q) \rightarrow (p \Rightarrow q)$

(T5.2) $L(\sim p) \rightarrow (p \Rightarrow q)$

The first may be established as follows: Assume that $L(q)$ is given. Note then that q can serve as the r satisfying the relationship: $L(r)$ & $([p \,\&\, r] \rightarrow q)$. The proof of (T5.2) proceeds in exactly the same fashion except that $\sim p$ serves as the required r.

A third somewhat similar relationship is represented by:

(T5.3) $(p \Rightarrow \sim p) \rightarrow \sim L(p)$

This may be established as follows

Assume (1) $p \Rightarrow \sim p$

(2) $(\exists r)[L(r) \,\&\, ([p \,\&\, r] \rightarrow \sim p)]$ From (1) by (D1.1).

(3) $\sim L(p)$ From (2) by (L5).

(4) $(p \Rightarrow \sim p) \rightarrow \sim L(p)$ From (1)—(3) by (R3.1).

A fourth relationship linking '$\Rightarrow$' with 'L' is:

(T5.4) $(p \Rightarrow q) \rightarrow L(p \supset q)$

This theorem is established as follows:

Assume (1) $p \Rightarrow q$

(2) $(\exists r)[L(r) \,\&\, ([r \,\&\, p] \rightarrow q)]$ From (1) by (D1.1).

(3) $(\exists r)[L(r) \,\&\, (r \rightarrow [p \supset q])]$ From (2) by (I2).

(4) $L(p \supset q)$ From (3) by (L4).

(5) $(p \Rightarrow q) \rightarrow L(p \supset q)$ From (1)—(4) by (R3.1).

[5] For this system see C. I. Lewis, and C. H. Langford, *Symbolic Logic* (N.Y., 1932), p. 500. The system S3 is governed by the eight axioms A1–A8 listed on p. 493. Here (T4.7), (T4.5), (T4.3), (T4.8) and (T4.2) correspond to A1–A5 respectively. Further (T4.9) corresponds to A7.

In consequence of (T5.4), we may say that a plausible implication statement '$p \Rightarrow q$' assures that 'if p, then q' is *likely* in the sense of the likelihood modality (subject to our construing "if-then" as material implication). This result further justifies us in construing the implication-relationship '$\Rightarrow$' given by (D1.1) as a formal counterpart of the informal notion of "plausible implication".

It warrants remark that '$L(p \supset q)$' is a very different statement from '$p \supset L(q)$'. For assume that 'p' and 'q' are such that '$p \,\&\, L(\sim p) \,\&\, \sim L(q)$' is true (as is perfectly possible, cf. section III above). Then '$p \supset L(q)$' is false, since its antecedent is true, but its consequent false. However '$L(p \supset q)$' or equivalently '$L(\sim p \vee q)$' is true, since '$L(\sim p)$' is true (by (L4)). This circumstance further illustrates the well-known fact that it is difficult to capture certain logical distinctions in ordinary language. For it would seem that both '$L(p \supset q)$' and '$p \supset L(q)$' correspond perfectly to the phraseology 'If p, then quite likely q'.

It is noteworthy that the converse of T5.4 may also obtain:[6]

$$(T^*5.5) \quad L(p \supset q) \rightarrow (p \Rightarrow q)$$

This is readily established by making use of the fact that, given the antecedent of (T*5.5), we have that $p \supset q$ can serve as the r satisfying the relationship: $L(r) \,\&\, ([p \,\&\, r] \rightarrow q)$. This relationship, in conjunction with T5.4, establishes the equivalence of '$p \Rightarrow q$' and '$L(p \supset q)$' within that (restricted) class of "mixed" systems of strict and material implication in which we have the *modus ponens* principle:

$$[p \,\&\, (p \supset q)] \rightarrow q.$$

This principle is *not* a universally acceptable thesis for "mixed" systems of strict and material implication. In particular, it must be excluded from systems in which the law of exportation obtains in the form: $[(p \,\&\, q) \rightarrow r] \rightarrow [p \rightarrow (q \rightarrow r)]$. For the conjunction of these two principles leads to $(p \supset q) \rightarrow (p \rightarrow q)$, with the result that all distinction between $\supset$ and $\rightarrow$ is abolished, and the two relationships become identified. This same result would be forthcoming if a deduction theorem governs strict implication, so that we have: If A and B jointly entail C then A entails $B \rightarrow C$. In consequence, it is not uncommon for "mixed" systems of strict and material implication to exclude the *modus ponens* principle: $[p \,\&\, (p \supset q)] \rightarrow q$.

VI

The considerations adduced thus far tend to suggest that the plausible implication relationship '$\Rightarrow$' bears a substantial degree of similarity to strict (and to material) implication. But we come now to a very marked point of difference: the failure of '$\Rightarrow$' to be governed by various forms of the *modus ponens* principle.

[6] This result was pointed out to the writer by the Acting Editor,

First of all, it is clear that we do *not* have:

(N6.1) $[p \,\&\, (p \Rightarrow q)] \rightarrow q$.

For assume that $p \,\&\, \sim q \,\&\, L(q)$ is the case. Then the consequent of (N6.1) is false; whereas its antecedent is true, since p is the case, and since q can serve as the r such that: $L(r) \,\&\, ([p \,\&\, r] \rightarrow q)$.

Secondly, we do not have:

(N6.2) $[p \,\&\, (p \Rightarrow q)] \rightarrow L(q)$.

For assume that $p \,\&\, L(\sim p) \,\&\, \sim L(q)$ is the case. Then the consequent of (N6.2) is false, whereas its antecedent is true, since p is the case, and since $\sim p$ can serve as the r such that: $L(r) \,\&\, ([p \,\&\, r] \rightarrow q)$.

This failure of the *modus ponens* principle to characterize '$\Rightarrow$' has the consequence of rendering this a mode of implication whose logical force is very weak.

Continuing in the direction of thought raised in the consideration of (N6.1) and (N6.2), we may ask ourselves what (if anything) follows from: $L(p) \,\&\, (p \Rightarrow q)$? Here the assumption '$L(p) \,\&\, \sim q \,\&\, L(q)$' suffices to show the untenability of:

(N6.3) $[L(p) \,\&\, (p \Rightarrow q)] \rightarrow q$.

For the consequent is false, given this assumption, whereas the antecedent is true, since L(p) is the case, and since q can serve as the r such that: $L(r) \,\&\, ([p \,\&\, r] \rightarrow q)$.

Further, it is also possible to demonstrate the untenability of

(N6.4) $[L(p) \,\&\, (p \Rightarrow q)] \rightarrow L(q)$.

However, the establishment of this point is a somewhat more complex matter, for which we shall employ the machinery to be built up in the next section. But first we shall make one further addition to our list of invalid principles:

(N6.5) $(p \Rightarrow \sim p) \rightarrow \sim p$.

The untenability of (N6.5) may be shown by assuming that $p \,\&\, L(\sim p)$ is the case; for then its consequent is false, whereas its antecedent is true, since $\sim p$ can serve as the r such that $L(r) \,\&\, ([p \,\&\, r] \rightarrow \sim p)$.

The failure of the group (N6.1)—(N6.5) to yield any valid principles for '$\Rightarrow$' provides a clear illustration of the logical weakness of the linkage provided by this mode of implication.

It ought, however, to be noted that we do have the one *modus-ponens*-type principle,

(T6.1) $[N(p) \,\&\, (p \Rightarrow q)] \rightarrow L(q)$

where 'N' represents the propositional modality of necessity, '$N(p)$' being equivalent with '$\sim p \rightarrow p$'. (This theorem is a consequence of (T5.4).) This result may be contrasted with an analogue, namely,

(L5) $[L(p) \,\&\, (p \rightarrow q)] \rightarrow L(q)$.

In going from (L5) to (T6.1), we note that the increased strength in the move from '$L(p)$' to '$N(p)$' suffices to counterbalance, as it were, the increased weakness in going from '$p \rightarrow q$' to '$p \Rightarrow q$'; so that the

greater weakness of the implicational linkage is offset by the greater assurance of the antecedent.

VII

A model for discussing the applicability of these regarding the logical relationships being studied will prove useful. It is based upon the notion of *state-descriptions*, as familiar from their extensive role in the writings of Carnap.[7] In particular, we shall assume that these are just five possible states: S_1, S_2, S_3, S_4, and S_5. Let R(p), to be called the *range* of p, be that set of states for which p is the case. We note that the following relationships obtain:

(R1) $R(p \,\&\, q) = R(p) \cap R(q)$

(R2) $R(p \vee q) = R(p) \cup R(q)$

(R3) $R(\sim p)$ = the complement of R(p) in the set $\{S_1, S_2, S_3, S_4, S_5\}$

We may further remark that two other relationships hold:

(R4) ' $p \rightarrow q$ ' is true if, and only if $R(p) \subset R(q)$.

(R5) ' L(p) ' is true if, and only if R(p) contains three or more members.

It should be observed that all of the rules of our modal logic, and thus all of the valid rules for '$\Rightarrow$', are satisfied in this model.

VIII

Let us now return to the question of the validity of:

(N6.4) $[L(p) \,\&\, (p \Rightarrow q)] \rightarrow L(q)$.

The untenability of this principle is now readily shown by a counter-example. Let ' p ', ' q ', and ' r ', be statements such that:

$R(p) = \{S_1, S_2, S_3\}$

$R(q) = \{S_1, S_2\}$

$R(r) = \{S_1, S_2, S_4\}$

Notice that we now have: (i) L(p), (ii) L(r) & ([p & r]$\rightarrow$q), and (iii) $\sim$L(q). This demonstrates the failure of (N6.4) to obtain as a general principle.

IX

A further manifestation of the logical weakness of plausible implication is its failure to obey the transitivity law:

(N9.1) $[(p \Rightarrow q) \,\&\, (q \Rightarrow r)] \rightarrow (p \Rightarrow r)$.

To demonstrate that (N9.1) is *not* a valid rule, we shall assume that ' p ', ' q ', ' r ', ' s ' and ' t ' are such that:

$R(p) = \{S_1, S_2, S_4, S_5\}$

$R(q) = \{S_1, S_2, S_3\}$

[7] A useful survey of the relevant ideas is presented in §18 of Carnap's *Logical Foundations of Probability* (Chicago, 1950).

$R(r) = \{S_3\}$
$R(s) = \{S_1, S_2, S_3\}$
$R(t) = \{S_3, S_4, S_5\}$

Now note that (i) L(s) & ([p & s]→q) obtains under our model, as does (ii) L(t) & ([q & t]→r); so that the antecedent of (N9.1) is satisfied in this interpretation. However there is no subset of $\{S_1, S_2, S_3, S_4, S_5\}$, having three or more members, whose intersection with R(p) is included in R(r), so that the denial of the consequent must obtain in the interpretation. As a result, (N9.1) must be excluded from our theory of the relationship '⇒'.

It should be observed, however, that the relationship of plausible implication is governed by the following two transitivity laws,[8] which are analogous to (N9.1), but have a strengthened antecedent:

(T9.1) [(p→q) & (q⇒r)]→(p⇒r)
(T9.2) [(p⇒q) & (q→r)]→(p⇒r)

The proofs of these two theorems are obvious.

X

Our discussion may now be brought to its conclusion with a brief word of retrospect. We have examined some basic characteristics of the logical theory of the relationship designated as *plausible implication.* We have seen that, while this relationship is in various ways analogous to the familiar connectives of *strict* and *material* implication, it is fundamentally weaker than both in the crucial respect that it is neither governed by a transitivity condition nor subject to a *modus ponens* rule (see the discussion of (N9.1) and of (N6.1) and its cognates). Furthermore, this relationship '⇒' is such that from 'p⇒~p', we cannot infer '~p', but only '~L(p)' (see (T5.3) and (N6.5)). In this respect too the relationship of plausible implication is logically "weaker" than strict or material implication.

Despite such weaknesses in its logical force, plausible implication would seem to be a notion which possesses some degree of interest, if only because it represents a very "natural" kind of logical linkage between statements. In fact, it is just through these points of weakness that plausible implication qualifies as an essentially *inductive* implication relationship. Thus the study of implication-relations of this kind may be recommended as a step towards the formulation of a logical theory of inductive modes of reasoning, a project which, though relatively undeveloped, is surely not without claims to importance and interest.

Lehigh University
Bethlehem, Pennsylvania

[8] This was pointed out to the writer by the Acting Editor.

WITTGENSTEIN AND LOGICAL COMPULSION

By CHARLES S. CHIHARA

IN " The Hardness of the Logical Must ",[1] Edward J. Nell takes Wittgenstein to be claiming in *Remarks on the Foundations of Mathematics*, I, 113, 115, that someone might reject outright our ideas of what constitutes identity, difference, and correctness of inference, and simply refuse to play " the language-game of logic " as we know it. He then goes on to criticize Wittgenstein, arguing that if a person did not see identity and difference as we do, a conversation of the sort presented by Wittgenstein could not take place, " for such a person could not *recognize* disagreement ".[2]

But what does it mean to say of someone ' He does not see identity and difference as we do '? This question is extremely difficult to answer from the information Nell gives us. We are told not only that the person could not recognize disagreement, but also that he would be unable to " remain either human or alive ".[3] I am afraid these hints only make the expression more puzzling.

Did Wittgenstein think that there could be a society with entirely different concepts of " identity " and " difference " from us? What would that be like? In the *Investigations*,[4] he wrote: " The common behaviour of mankind is the system of reference by means of which we interpret an unknown language." But how could we compare a completely different notion of " sameness " with ours? What could we point to as " their concept of sameness "? Why call it " sameness "? We seem to run into insuperable difficulties if we take this line, and the issues become exceedingly murky.

However, I do think Wittgenstein is saying something intelligible in these sections, which at least merits consideration. Many of us are inclined to think we are logically compelled to do certain things, without being at all clear about the nature of this compulsion. If we begin to develop a sequence according to a particular rule, it does seem that we are compelled by the rule to put down a certain number at a particular place in the sequence. As against this, Wittgenstein writes: " —Compelled? After all I can presumably go as I choose! " Here, he is simply pointing out that, in one sense at least, we certainly are not compelled to write down any particular number at a particular place in the sequence, since we are not compelled to write anything: one might just put down one's pencil.

[1] ANALYSIS, 21.3, January, 1961.
[2] *Ibid.*, p. 71.
[3] *Ibid.*, p. 70.
[4] I, 206.

A natural reply to this, which Wittgenstein considers, is: " —But if you want to remain in accord with the rules you *must* go this way." Still, if we are not compelled to take this rather than that as being in accord with the rule, how can we be said to be compelled to get this at all? But if I am compelled to take this and only this as being in accord with the rule, what compels me? Does the meaning of the rule compel? How can the meaning compel me to take this as the meaning? Where is the compulsion to take this as being in accord with the rule? It now seems somewhat difficult to specify just what the compulsion is.

Suppose, in answer to the assertion ' If you want to remain in accord with the rules you *must* go this way ', someone replies as suggested by Wittgenstein: ' —Not at all, I call *this* " accord ".' Perhaps we would retort ' You can *say* that is in accord with the rules; you can say anything you please; but you surely can't *think* it '. Wittgenstein considers such a reply in section 116 where he writes:

> . . . I am only saying that that means, not: try as he may he can't think, it, but: it is for us an essential part of ' thinking ' that—in talking, writing, etc.—he makes *this sort* of transitions. And I say further that the line between what we include in ' thinking ' and what we no longer include in ' thinking ' is no more a hard and fast one than the line between what is still and what is no longer called " regularity ".

Is the real question here what is the meaning of the expression ' being in accord '? For Wittgenstein, such terms as ' accord ', ' agreement ' and ' same ' are intertwined with ' rule '—" they are cousins ".[5]

The point Wittgenstein is making can perhaps be seen in terms of the following example. Suppose you write down the sequence 3, 6, 9, and say to a person: ' Now continue the sequence doing the same thing as I did in obtaining each term from its preceding term.' Imagine however that the person puts down 13, 17, 21. You immediately interrupt him saying: ' That's not right—you are not doing the same thing as I did.' If, after checking what had been done, he asks ' But what is " doing the same thing "? ', it would probably do no good to explain the meaning of the word ' same '. Might there not be various things he could do which could be correctly described as " doing the same thing as you "? And could he not also say ' But I am doing the same thing as you did '? That is, could he not reply like a rational person, and yet not be playing your game? (Here, I take it, not even Mr. Nell would be tempted to interpret ' game ' as referring to some " language-game of logic ").

And suppose you than say ' Look—just continue in the obvious way '. Would it not be possible for someone to take it as obvious that since one began the sequence from 0 by adding 3 three times to get the first

[5] *Philosophical Investigations*, I, 224.

three terms, one should continue the sequence by adding 4 four times, 5 five times, 6 six times, and so on? Of course, a child who proceeded in this manner on an I.Q. test would probably be marked down for it. But if children were only taught to develop sequences of natural numbers according to rules like: 'add 1 once, add 2 twice, and 3 thrice, etc.', we would consider continuations of 3, 6, 9, in the above manner to be quite proper.

Imagine there are people, living in a far-away land called Myo, who consider sequences of the above sort to be fundamental and primary. If you asked one of the Myonese to continue the sequence 3, 6, 9, by doing the same thing as you in deriving each term from its preceding term, might he not think that you were following the procedure which we would describe as 'adding 3 three times, 4 four times, etc.'?

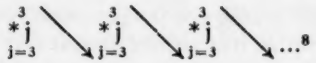
Mr. Nell argues: "When one proposes a new or different interpretation, or challenges that of another, one must justify oneself, one must present reasons for one's views. . . ."[6] Now what reasons can we give for saying 3, 6, 9, is obviously to be continued 12, 15, 18, . . .? I am tempted to say 'It strikes me that way', but this does not *justify* my interpretation. I suppose if I had to give a reason, I would say to the Myonese something like this: 'As you continue the sequence, you must do something *different*, for the operation of adding 4 was not performed in deriving the second term from the first, or the third from the second. Furthermore, my way of continuing the sequence is clearly the simpler.'

Suppose however that the Myonese, for some reason, very easily learn the operation which we would describe as adding 1 once, adding 2 twice, adding 3 three times, etc. They use the notation "$\overset{n}{*}_{j=a}\,j$" for operations of this sort where we would define $\overset{n}{*}_{j=a}\,j$ as the operation of adding a a times, adding (a+1) (a+1) times, (a+2) (a+2) times, ..., n n times. The Myonese use this operation to define sequences of natural numbers, it being understood that one obtains the first term in the sequence by adding to 0 and that successive additions yield successive terms.[7]

Now though *j operations come easily to the Myonese, the operation which we describe as 'adding 3 continually' is extremely difficult for them to master: they have a strong tendency to stop adding 3 after the third time. The operation which we describe as 'adding 3 continually' the Myonese see as:

[6] Nell, *op. cit.*, p. 68.

[7] The Myonese "*j" notation is obviously similar to our "Σj" notation. Of course '$\overset{n}{\Sigma}_{j=a}\,j$' is our symbol for the *sum* a+(a+1)+(a+2)+...+n, whereas $\overset{n}{*}_{j=a}\,j$ stands for either an operation or a sequence.



When it comes to developing sequences of natural numbers, *j operations are considered fundamental. The Myonese would not think they were doing anything different in continuing the *same* *j operation, which (they thought) we were performing in putting down 3, 6, 9; and they would not dream that we expected them to perform a very "complicated" operation in which one breaks off the *j operation for j=3 and begins a "new" *j operation for j=3.

We are inclined to think situations of the above sort could not arise if a precise mathematical definition of the sequence were given. If we had precisely defined the rule we were following in developing the first three terms of the sequence, there could be no question, it would seem, as to what "doing the same thing" would be. Wittgenstein, rightly or wrongly, assimilated even cases involving precise definitions to the above type of case; and I think it is in connection with such thoughts that Wittgenstein considers the example of the person being taught to develop the sequence of even natural numbers.[9] In explaining the sequence to someone, we might give him the formula '2n: n=1, 2, 3, . . .' But to be of any use, the pupil would have to know how to apply the formula. And how would we teach him that? Well, we might explain the formula in terms of the recursive formula '$A_{n+1}=A_n+2$: n=1, 2, 3, . . .' But how would we teach him to apply the latter? Ultimately we come to a situation very much like the one we started with: we give him various examples of applying the formula; we start him off and encourage him to continue; we correct his mistakes, etc. Now isn't it possible that after apparently learning how to apply the formulas, the pupil may suddenly, after reaching, say, 1000, begin to do what *we* would describe as 'adding 4 to the preceding term'? And if we say 'Look—you are making a different application of the formula' is it not possible that he might sincerely reply, 'Different?—But this surely *isn't* different!'? This will probably seem incredible to some people. One may want to ask 'How could anyone think that?' And here, Wittgenstein's comment in section 116, which Mr. Nell so badly misconstrues,[10] does make sense, for there is "something in saying: he can't *think* it. One is trying e.g. to say: he can't fill it with personal

[8] The reader may wonder why the Myonese use the symbol "↘" since it may seem easier to describe the operation as $\overset{3}{*}j,\ \overset{3}{*}j,\ \overset{3}{*}j,\ \ldots$ The Myonese, however, reserve "$\overset{3}{*}j,\ \overset{3}{*}j,\ \overset{3}{*}j,\ \ldots$" for the infinite sequence of finite sequences: (3, 6, 9), (3, 6, 9), . . . The arrow also serves to remind the Myonese that they are not to continue the *j operation beyond j = 3 but must begin a "new" *j operation for j = 3.

[9] *Philosophical Investigations*, I, 185.

[10] Nell, *op. cit.*, p. 72.

content; he can't really *go along with it*—personally, with his intelligence." It strikes us as incredible that anyone should take the explanations in such an odd way; that way of taking the explanations makes no sense to us.[11] "It is like when one says: this sequence of notes makes no sense, I can't sing it with expression. I cannot *respond* to it. Or, what comes to the same thing here: I don't respond to it."[12]

Of course, cases of such deviant reactions to explanations and training are necessarily rare; but they are not impossible. And a person who frequently took our utterances, signs, explanations and the like so differently from the rest of us would most certainly "get into conflict, e.g. with society; and also with other practical consequences".[13]

That these considerations by no means settle the issue Wittgenstein certainly realized (as can be seen from the fact that he returned again and again to this puzzling notion of logical compulsion). Still, what he says in these sections of the *Remarks* throws much light on many of his paradoxical statements about the nature of logic and mathematics. And whether Wittgenstein is right or wrong, I think what he says here is worth serious examination and should not be set aside as the utter nonsense Mr. Nell would have us believe it to be.

University of Oxford

PARADOXES OF SELF-DECEPTION

By JOHN CANFIELD *and* PATRICK MCNALLY

RECENTLY Professor Raphael Demos has pointed out and offered a solution to a paradox involved in the notion of self-deception.[1] There is a strong sense of deceiving such that A deceives B only if A intends to lie to B, A succeeds in making B believe the lie, and A knows that what he tells B is false (p. 588). When we wish to say in this sense that a person deceives himself, we meet a paradox. As Demos writes, "When B lies to himself he comes to believe what he knows to be false; to accept this as a description of a fact is to admit a violation

[11] It is not difficult, however, to fill in the example, as in the previous case, in such a way that this way of taking the explanations would not strike us as being so unintelligible.

[12] *Remarks on the Foundations of Mathematics*, I, 116.

[13] *Ibid.*

[1] Raphael Demos, "Lying to Oneself". *Journal of Philosophy*, Vol. LVII, No. 18, September 1, 1960.

of the law of contradiction " (p. 591). An example of a case of self-deception which generates the paradox is this (p. 589): A woman knows her son is bad, but fools herself into believing he is good. She then believes he is good, but at the same time, in some way or other, knows and believes that he is bad. We thus say, ' She believes her son is wholly good and she believes he is wholly bad '. The paradox is, then, that although such cases of self-deception exist, our description of them seems to be self-contradictory. In this paper we shall examine Demos' statement of the problem and his solution, and we shall try to show that there remain unsolved two paradoxes of self-deception whose form can be exhibited in a fairly unambiguous way.

The solution Demos offers to the paradox he has elicited is built on the concept of *noticing* or *attending to* (p. 594). We can have a headache and yet not notice that we have it because, for example, we are engrossed in a movie. This is distinguished from the case in which our headache disappears when we take aspirin. Similarly, we can have beliefs which are in some way ' before the mind ' and yet not notice or attend to them. In self-deception this is what happens. We attend to the belief not-p, while simultaneously we have, but do not notice, the belief p. Thus Demos writes:

> The person who lies to himself fails to notice or ignores what he knows to be the case. Such an analysis ' saves ' the phenomena while at the same time conforming to the requirements of the law of contradiction. For, indeed, we are saying that the person who lies to himself believes both p and not-p, and is capable of doing so because he is distracted from the former (p. 594).

An examination of Demos' position must focus first on his statement of the paradox. For he has failed to make explicit a distinction between two different types of description of self-deception. We describe cases of self-deception in this way:

(I) Jones believes p and Jones believes not-p.

Demos takes (I) to refer to an instance of self-deception. But he also writes, for example, in describing a rival solution to the paradox, " In self-deception, believing p and disbelieving p occur at different and successive times" (p. 591), thus indicating that he tends to amalgamate (I) with:

(II) Jones believes p and Jones does not believe p.

(I) and (II), however, must not be taken as synonymous or as nearly synonymous. Ordinary usage seems to warrant the inference from " He believes not-p " to " He does not believe (or disbelieves) p." Thus (II) can be inferred from (I). The inference cannot go the other way, however. If Jones does not believe that a god exists, for example, we cannot infer

that he believes that a god does not exist; this would be to infer that he is an atheist, whereas in fact he might be an agnostic. (I) and (II) differ in other respects. In (II), but not (I), the intentional object, i.e. what is said to be believed or disbelieved, is one and the same. And only (II) has even the appearance of being self-contradictory. The paradox involved in the notion of self-deception will have a different form corresponding to these two different descriptions of self-deception. We shall examine the resulting two paradoxes of self-deception and test the adequacy of Demos' solution against each.

Because p and not-p are not truth-functional components of (I), (I) is not self-contradictory and in fact will often be true. Nevertheless many instances of (I) will have a paradoxical character. They will not be self-contradictory, but logically odd. The sentence ' It is raining but I don't believe it ', is logically odd and an *informal* contradiction. Similarly, the sentence ' She believes that her son is wholly good and that he is wholly bad ' is logically odd and an *informal* contradiction. Yet we may suppose that this sentence refers to a known fact.

Demos would avoid the logical oddity by supplementing the description with information about what the woman is noticing. We now say ' She believes, and notices the belief, that her son is wholly good, and believes, and does not notice the belief, that he is wholly bad.' The force of this solution is heightened if we treat the state of believing *and* noticing as a new intentional state ' believing-and-noticing ' and contrast it with the state of ' believing-and-ignoring '. The corresponding sentence made with the new intentional verbs will have no trace of logical oddity.

Although this solution seems adequate for some instances of self-deception, it is incomplete. This is obvious in the case of dispositional beliefs. When the woman is asleep, of example, it is still true that she believes her son is wholly good, even though she knows, and hence believes, that he is wholly bad. Thus our description of the sleeping woman's beliefs is at once true and an informal contradiction. Clearly the solution Demos offers is irrelevant here, for the sleeping woman can notice neither belief. An overlapping set of cases which cannot be treated as Demos suggests is made up of those where one of the beliefs is an unconscious or repressed one. It makes no sense to speak of failing to notice or attend to a repressed belief, when ' notice ' and ' attend ' are explicated by analogy to our noticing or attending to a present headache from which we have been momentarily distracted. To bring a repressed belief to consciousness requires more than such a mere turning of our attention to it. Thus under (I) there are paradoxical cases for which Demos' solution will not work.

When we turn to descriptions of type (II), we see that the paradox here also involves an informal, rather than an appearance of a formal,

contradiction. Here Demos' solution is inadequate for all instances of the paradox. Consider a person who has fooled himself into believing he is a great artist. We say he behaves as if he were a great artist and that in fact he actually believes he is. On the other hand, we say that deep inside or " in some corner of his mind " he really does not believe he is a great artist. Since he really does not believe it, he does not believe it. Thus from a description of an actual state of affairs we derive:

(IIa) He believes he is a great artist and he does not believe he is a great artist.

There are two ways of interpreting (IIa), one of which yields a formal contradiction, the other not. Idiomatically, ' does not believe ' is equivalent to ' disbelieves '. Thus (IIa) can be rewritten:

(IIb) He believes that he is a great artist and he disbelieves that he is a great artist.[2]

On the other hand, if (IIa) is interpreted as self-contradictory it must be rendered as:

(III) He believes he is a great artist and it is false that he believes he is a great artist,
where the second conjunct is simply the negation of the first and means that he has no belief that he is a great artist.

It is clear, however, that (III) does not fit the case described. The terms ' deep inside ' and ' in some corner of his mind ' indicate that in some way the person actually disbelieves in his ability (or believes in his inability, from which the disbelief in his ability follows). And as the case is described, it is false that he has no belief in his ability.

Not (III) but (IIb) then is a correct description of the artist's state of belief. (IIb) is obviously not self-contradictory. If there is a paradox involved in this case, then, it is again that our description of an actual state of affairs appears logically odd, or appears to be an informal contradiction.

It is easily seen that the solution Demos constructs in terms of noticing and ignoring has no relevance to cases such as (IIa). As has been noted, in such cases the intentional object, what is believed or disbelieved, is identical in both conjuncts of the logically odd sentence. Thus if we say in the first conjunct that the person notices the belief, we must also say in the second that he notices what is disbelieved. The sentence formed by using the new intentional verbs ' believe-and-notice ', ' believe-and-ignore ', and so on, will then read: ' He believes-and-notices that he is a great artist, and (does not believe)-and-notices

[2] We are indebted to the Editor for pointing out to us both the idiomatic equivalence between (IIa) and (IIb), and also the fact that in this case (IIa) must not be interpreted as a formal contradiction such as (III).

(disbelieves-and-notices) that he is a great artist '. Plainly if (II) or (IIa) is logically odd, so is the later sentence. We can transport the paradox-saving phrase ' disbelieves-and-ignores ' into the second conjunct of the sentence only at the cost of asserting the much more paradoxical proposition that a person can simultaneously notice and ignore the same intentional object.

Demos sees descriptions of self-deception as involving a seeming formal contradiction. The cases we have examined, however, involve only informal contradictions. In a case of self-deception a formal contradiction would look something like (III). (III) is only paradoxical, however, if each of its conjuncts is either a part of, or entailed by part of, a true description of an actual state of affairs. Obviously it would not be paradoxical if the first of these conjuncts were merely *confirmed* by part of the evidence, and the second and contradictory conjunct confirmed by another part (or indeed the same part) of the evidence. In the fable, the first blind man, feeling the elephant's tail, says the animal is like a snake. The second blind man, feeling the elephant's side, says no, it is like a wall. Similarly, it is unparadoxical if part of our evidence, say a person's dreams and verbal slips, confirms that the person has no belief that p, while another part, his assertions, confirms that he believes p. And clearly it will never be the case (it is logically impossible) that part of the accurately gathered evidence contradicts, or entails a contradiction, of another part. If a formal contradiction seems to be involved in self-deception, then, it is because ' He believes p and not-p ' has been mistaken for a formal contradiction, or because ' He believes p and he does not believe p ' has been mistakenly interpreted as ' He believes p and he has no belief that p ' rather than as ' He believes p and he disbelieves p '.

In summary, we have distinguished two different paradoxes of self-deception. In the first class of paradoxical cases, we argued that Demos' solution is inadequate for a large number of instances, in particular for those instances where the beliefs in question are dispositional ones. In the second class of instances we argued that Demos' solution is inadequate because of the identity of the intentional objects in both conjuncts of the belief statements. The paradoxes have been seen to be of the same kind. In both, a description of an actual state of self-deception is, *prima facie*, logically odd or an informal contradiction.

Brown University; University of Göttingen